

COMPUTER PROGRAMING 2

WEB SCRAPING

MODUL 15

2023

DAFTAR ISI

DAFTAR ISI	2
WEB SCRIPING	3
A. Pengertian Web Scraping	3
B. Cara Kerja Web Scraping	3
C. Mengapa Web Scraping penting?	3
D. Fungsi Web Scraping	4
E. Tutorial Web Scraping	4
REFERENSI	9

WEB SCRIPING

Teknik pengumpulan data sekunder memang mulai banyak dilirik. Selain karena disebabkan oleh pengambilan data primer yang memakan lebih banyak waktu, biaya, dan tenaga, data sekunder juga dapat digunakan untuk menguji dan memperbaiki hasil penelitian sebelumnya. Ada banyak website resmi yang menyediakan data sekunder dan membolehkan masyarakat umum untuk menggunakannya, seperti BPS.

Sama seperti proses pengumpulan data primer yang memiliki beberapa teknik, data sekunder juga demikian. Salah satu teknik yang bisa digunakan adalah web scraping. Teknik ini akan sangat berguna jika data yang ingin diambil berjumlah sangat banyak atau bisa dikatakan sebagai big data. Dalam artikel ini kita akan membahas mengenai teknik web scraping lebih lanjut, mulai dari pengertian, cara kerja, dan alasan mengapa web scraping menjadi cukup penting untuk dikuasai di era ini.

A. Pengertian Web Scraping

Web scraping merupakan proses pengumpulan data dan informasi yang ada dalam sebuah website secara otomatis dan spesifik [1]. Dikatakan spesifik karena data yang diambil hanya di bagian tertentu saja sesuai dengan kebutuhan. Umumnya halaman web yang bisa dilakukan scraping merupakan halaman web yang dibuat dalam bahasa markup XHTML atau HTML. Untuk melakukan web scraping ini bisa menggunakan bahasa pemrograman Python, namun ada juga add-ins di google chrome yang bisa di instal untuk melakukan scraping.

B. Cara Kerja Web Scraping

Cara kerja dari web scraping dapat dibagi menjadi empat langkah, yang pertama adalah membuat program yang akan digunakan untuk mempelajari dokumen HTML yang ada di website yang dijadikan tujuan untuk diambil datanya. Langkah kedua merupakan proses pembuatan program yang digunakan untuk mempelajari teknik navigasi pada website tersebut, sehingga kita dapat mengaplikasikannya pada aplikasi web scraper yang dibuat. Langkah selanjutnya adalah proses pengambilan data secara otomatis dari website yang telah ditentukan, langkah ini baru bisa dilakukan ketika langkah pertama dan kedua telah berhasil. Dan yang terakhir, menyimpan data yang didapatkan dalam database.

C. Mengapa Web Scraping penting?

Teknik web scraping dapat memudahkan pekerjaan manusia dalam proses mengumpulkan data. Kita tidak perlu lagi melakukan survey untuk hal-hal yang sudah ada di website, contohnya kepuasan pelanggan. Kita cukup men-scraping data yang ada di website tersebut. Perkembangan big data juga sangat mendukung penggunaan web scraping, karena semua data yang berukuran besar dan bervariasi tersebut dapat ditarik secara otomatis. Namun sayangnya,

untuk melakukan web scraping tidak semudah yang dibayangkan, dibutuhkan kemampuan tingkat tinggi di bidang pemrograman untuk melakukannya.

D. Fungsi Web Scraping

Perlu diketahui kalau fungsi utama web scrapping sebagai pengumpul data dan informasi [2]. Selain itu, ada fungsi lain dari web scrapping yang tidak kalah penting. Berikut adalah fungsi dari web scriping:

1. Monitoring harga di pasaran
Adanya web scrapping bisa memonitoring harga produk yang sama di pasaran. Jika mempunyai produk tertentu, maka bisa dengan mudah melihat perbandingan harganya dengan memakai web scrapping.
2. Analisa competitor
Bukan hanya bisa mengetahui harga produk yang serupa dari competitor. Adanya web scrapping juga bisa memberitahukan dengan dalam berbagai macam informasi perusahaan lain. Web scrapping akan melakukan analisa competitor. Data tersebut akan berguna dalam pengambilan keputusan dan merencanakan strategi yang tepat dalam berbisnis.
3. Mendapatkan informasi yang akurat dan actual
Anda dapat menggunakan beberapa tools dalam melakukan Web Scraping seperti Google Spreadsheet, Octoparse dan lainnya. Adanya web scrapping bisa mengetahui berbagai macam tren yang kini beredar di masyarakat. Selain itu juga bisa memantau berbagai macam informasi yang akurat dan terpercaya dari website dengan memakai web scrapping. Itulah pengertian dari web scrapping dan fungsinya. Web scraping adalah suatu metode yang memudahkan untuk menjalankan sebuah bisnis, terutama bisnis online. Selain itu, ada juga fungsi lainnya yang bisa mmebantu untuk menyukkseskan bisnis.

E. Tutorial Web Scraping

Untuk melakukan web scriping terdapat beberapa Langkah yang harus dilakukan diantaranya sebagai berikut:

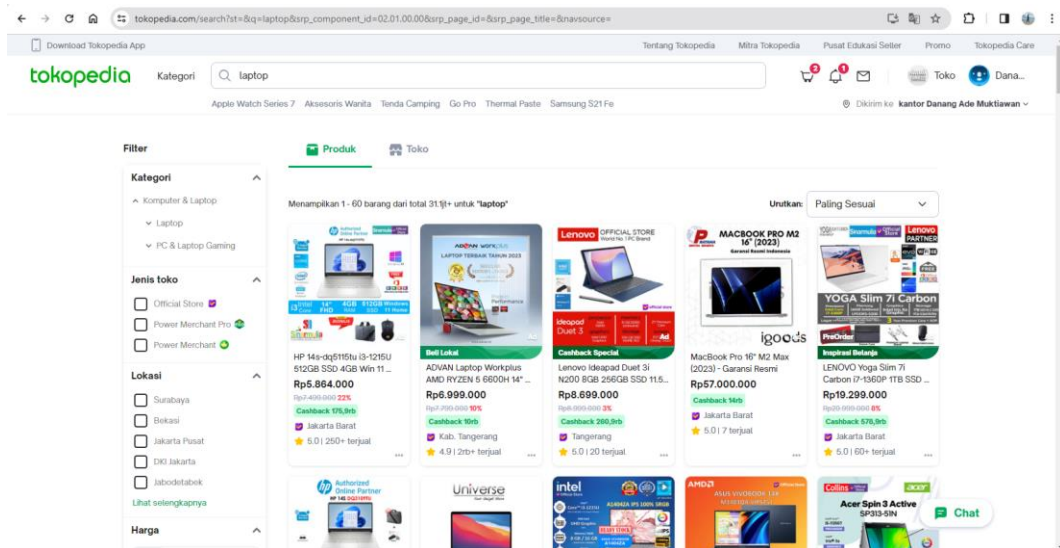
```
import time
from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions
as EC
from selenium.webdriver.common.by import By
```

Silahkan running program diatas. Jika pada saat running program terdapat eror kemungkinan besar yaitu perlu dilakukannya install library yang erorr. Cara untuk

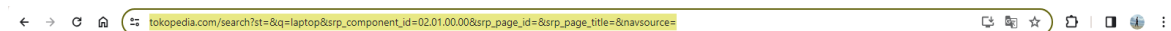
menginstall library pada Bahasa pemrograman python yaitu dengan melakukan intruksi berikut:

```
Pip install namalibrary
```

Selanjutnya yaitu menambahkan url yang akan kita lakukan scraping data. Untuk mengetahui Alamat url yang akan kita scraping yaitu dengan cara mengunjungi halaman (sebagai contoh yaitu "Tokopedia"). Silahkan kunjungi tokopedia.com kemudian cari barang yang akan kita scraping. Sebagai contoh kita akan melakukan scraping data penjualan laptop. Jadi pada kolom pencarian barang kita ketikkan laptop, kemudian kita copy link url yang ada pada browser seperti pada gambar berikut:



Kita cukup focus pada link url terlebih dahulu seperti gambar berikut:



Linknya yaitu :

```
https://www.tokopedia.com/search?st=&q=laptop&srp_component_id=02.01.00.00&srp_page_id=&srp_page_title=&navsource=
```

Jika link url sudah didapatkan selanjutnya yaitu menambahkan program berikut ini:

```
url=https://www.tokopedia.com/search?st=&q=laptop&srp_component_id=02.01.00.00&srp_page_id=&srp_page_title=&navsource="
driver = webdriver.Chrome()
driver.get(url)
```

Selanjutnya kita tambahkan script untuk menampilkan dan memparser program html sehingga nantinya akan lebih mudah dalam mencari data. Berikut adalah program yang akan ditambahkan:

```
soup = BeautifulSoup(driver.page_source, "html.parser")
print(soup)
```

Jika kita lihat hasilnya akan seperti gambar dibawah ini.

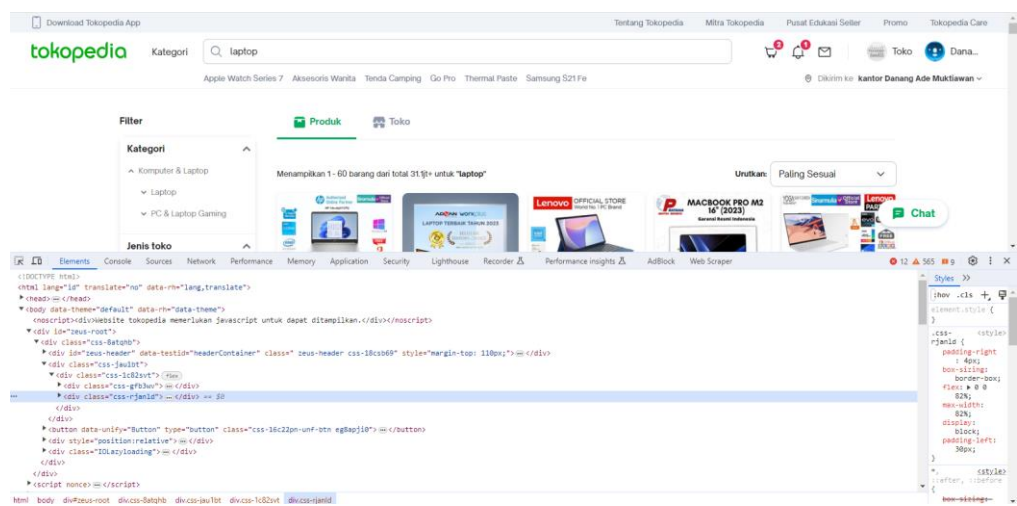


Link html pada website Tokopedia sudah kita dapatkan. Selanjutnya yaitu mencari class untuk mencari data yang diinginkan. Sebagai contoh yaitu nama produk dan harga. Berikut adalah program untuk memarsing data.

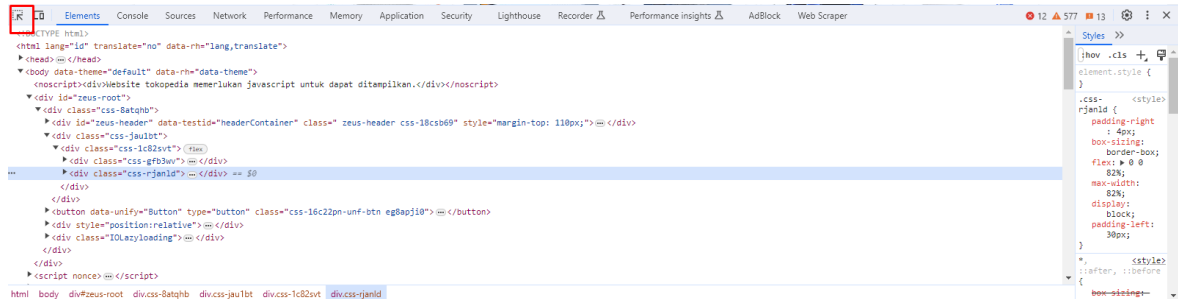
```
for item in soup.find_all('div', class_="css-1asz3by"):
    nama_produk = item.find("div", class_="prd_link-product-name css-3um8ox").text
    print(nama_produk)

time.sleep(3)
driver.close()
```

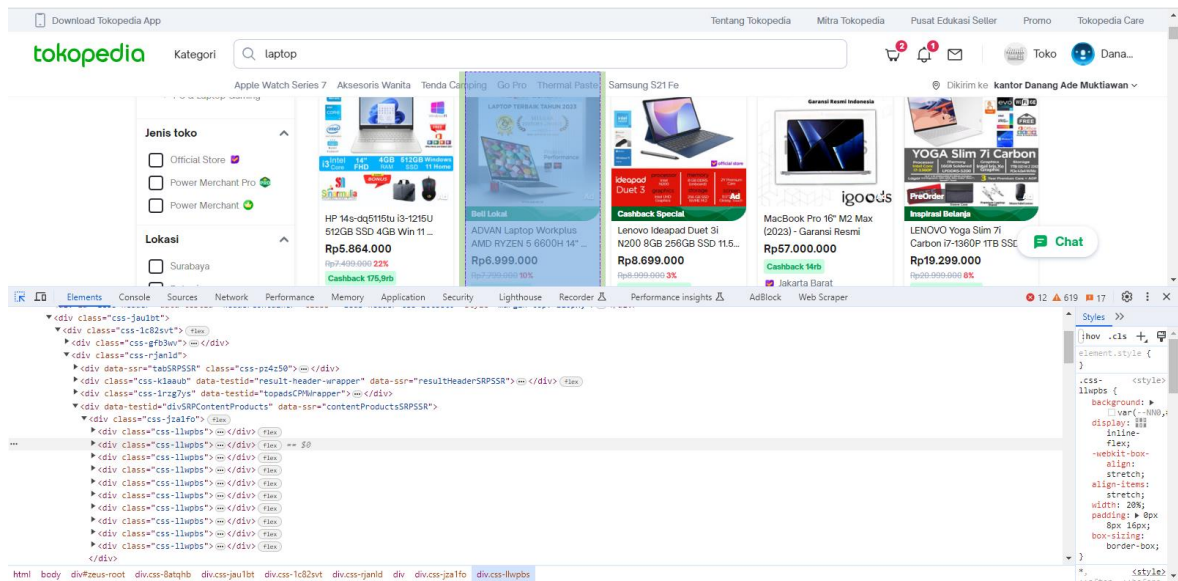
Untuk mencari class html, Langkah yang harus kita lakukan yaitu melakukan inspect element pada browser. Caranya yaitu dengan klik kanan sembarang kemudian pilih inspect dan hasilnya sebagai berikut:



Akan muncul jendela baru dibagian bawah. Untuk mempermudah dalam melakukan pencarian class yang lebih spesifik mari kita minta bantuan tools yang ada pada browsernya. Letaknya yaitu dibagian pojok kiri atas jendela baru yang muncul.



Fungsi dari tools tersebut yaitu untuk mempermudah dalam mencari element yang kita butuhkan. Cara penggunaannya yaitu dengan kita klik satu kali pada tools tersebut kemudian kita arahkan ke bagian yang akan kita acari. Sebagai contoh yaitu judul produk. Berikut adalah tampilannya pada penggunaan toolsnya.



Berdasarkan source code html yang ada, produk laptop berada pada class "css-1asz3by", dan untuk judul produk berada pada sub class "prd_link-product-name css-3um8ox" seperti pada gambar berikut:

```

<div class="css-1asz3by" flex>
  <a class="bcv3 info-content css-gwkt0u" href="https://ta.tokopedia.com/promo/v1/clicks/8a-xgVY2gmUEH_yfo_07b=23" title="ASUS VIVOBOOK 14 A416MAO FHD425 FHD426 - N4020/4GB/256GB/14&#34;/W11+OHS21"> flex == $0
    <div class="prd_link-product-name css-3um8ox" data-testid="spnSRPProdName">ASUS VIVOBOOK 14 A416MAO FHD425
      <div class=""></div>
      <div class="css-yaxhi2" data-productinfo="true"></div>
    </a>
    <div class="css-lw1xpyf"> flex
  
```

Dari program diatas hasil yang didapatkan yaitu sebagai berikut:

```
DevTools listening on ws://127.0.0.1:52681/devtools/browser/2a056965-a62e-4249-87ed-6c3d93ee443a
Laptop ADVAN Soulmate Celeron N4020 4GB/128GB (upgradable) 14" HD - BLUE TANPA TAS
ADVAN Notebook Laptop Workpro Intel I5 14'' FHD IPS 8GB 256GB Win 11 - Grey
ADVAN Laptop Workplus AMD RYZEN 5 6600H 14'' FHD IPS 16GB 512GB Win 11 - 512 GB
HP 14s-dq5115tu i3-1215U 512GB SSD 4GB Win 11 Home + OHS
Laptop Lenovo IdeaPad Slim 3i i3-1215U 512GB SSD 8GB Win11+OHS
Lenovo Legion Slim 5i i7-13700H RTX4060 512GB SSD 16GB Win11+OHS
ADVAN Laptop Workplus AMD RYZEN 5 6600H 14'' FHD IPS 16GB 512GB Win 11
MacBook Pro 16" M2 Max (2023) - Garansi Resmi
Apple Macbook Air M1 256 GB/13"/8GB - Garansi Resmi iBox
ASUS VIVOBOOK TERMURAH GEN12 A1405ZA - I3 1215U 16GB 512SSD 14" WUXGA
ASUS VIVOBOOK 14X M1403QA RYZEN 7 5800H 16GB 512SSD W11+OHS IPS WUXGA
MacBook Air 2022 M2 Chip 13" Inch 512GB 256GB RAM 8GB Apple IBOX
Acer Spin 3 Active SP313-51N i5-1135G7 8GB 512GB Iris Xe 13.3"FHD W10
Lenovo Ideapad Duet 3i N200 8GB 256GB SSD 11.5" 2K IPS 5MP Win11+OHS
```

Dada nama produk berhasil kita dapatkan. Untuk informasi yang lain silahkan dikembangkan ya..

=== To be Continue ===

REFERENSI

- [1] A. "DQLAB," Kenali Web Scraping, Salah Satu Teknik Pengumpulan Data Sekunder!, 23 2021. [Online]. Available: <https://dqlab.id/kenali-web-scraping-salah-satu-teknik-pengumpulan-data-sekunder>. [Accessed 8 1 2024].
- [2] M. "Telkom University," Web Scraping : Pengertian dan Fungsinya dalam Pengambilan Data, 24 12 2022. [Online]. Available: <https://it.telkomuniversity.ac.id/web-scraping-pengertian-dan-fungsinya-dalam-pengambilan-data/>. [Accessed 8 1 2024].